

Variation in Height Acceleration in the Fels Growth Data

J. O. Ramsay – McGill University

N. Altman – Cornell University

R. Darrell Bock – University of Chicago

January 8, 1993

The research reported in this paper was supported by grant APA0320 from the Natural Sciences and Engineering Research Council of Canada to J. O. Ramsay, grant HD-25031 to the University of Chicago, and grants HD-12252, HD-26971, HD-27063, and AG-08510 to Wright State University

Abstract

The Fels growth data record at half yearly intervals the heights of children from birth to adulthood, and are the basis for pediatrician's growth charts used throughout North America. Aspects of human growth are the subject of a large medical and statistical literature. This paper uses smoothing splines to study the variation in height acceleration. By use of a functional version of principal components analysis, we find that variation in the acceleration curve is essentially three-dimensional in nature. Evidence for a small growth spurt between the ages of six and eight, reported for data collected in Switzerland, is examined, and little support is found for the existence of this phenomenon in the Fels data.

Keywords:

functional data analysis, smoothing splines, principal components analysis, derivative estimation, nonparametric regression

1 The Analysis of Human Growth Data

The description of the increase in height with age among growing humans is a statistical problem with a long history (Jenss and Bayley, 1937; Pütter, 1920), and it continues to generate much research. The goals in the analysis of height data are varied: to offer a satisfactory model for individual growth curves, to describe particular features of growth curves, such as the onset and duration of the pubertal growth spurt, to define the normal range of variation in growth curves, and to identify unusual growth patterns.

Much of the early work on growth curves made use of standard tools in multivariate analysis: polynomial regression, principal components analysis, and structural equation modelling. Rao (1987) reviews and extends his own pioneering work with multivariate and polynomial models, and the discussion accompanying this paper offers spirited commentary on the appropriateness of these methods in various contexts. These approaches have tended to pay little or no attention to the time-dependent or functional nature of the data.

Preece and Baines (1978) have proposed specific parametric models for height, usually involving logistic functions. Bock and his co-workers (Bock and Thissen, 1980; Du Toit, 1992) developed the eight-parameter triphasic logistic model

$$h(t) = C_1/(1 + e^{-B_1 t}) + C_2/(1 + e^{A_2 - B_2 t}) + C_3/(1 + e^{A_3 - B_3 t}) . \quad (1)$$

Each of the parameters has a useful substantive interpretation for characterizing growth for individuals.

More recently nonparametric techniques such as kernel smoothing (Gasser, Kühler, Müller, Kneip, Largo, Molinari, and Prader, 1984), and spline functions (Largo, Gasser, Prader, Stützel, and Huber, 1978) are used to model the data. Subsequent interest and analysis focuses on the features of the actual function fitting an individual's data.

Gasser, et al (1984) have a series of papers (Gasser, Müller, Köhler, Prader, Largo and Molinari, 1985; Gasser, Kneip, Ziegler, Largo and Prader, 1990; Gasser, Kneip, Binding, Prader and Molinari, 1991; Gasser, Kneip, Ziegler, Largo, Molinari and Prader, 1991) showing the importance of the second derivative or acceleration in height in studying aspects of growth. Working primarily with variants of kernel smoothing, they have been able to demonstrate the existence of a local maximum in acceleration at around

seven years of age, which they call the mid-spurt (MS). Unlike the well-known pubertal growth spurt (PS), this feature shows near zero acceleration at the maximum, and thus consists in growth deceleration before and after a brief period of near linear growth. Its timing is also not gender-dependent and less variable than the PS. Their data come from the Zurich Longitudinal Growth Study, and involve other measures of growth as well as skeletal height, and in fact other measures such as shoulder width display the mid-spurt much more obviously.

This paper uses spline smoothing to study variation among growth curves, with a view to defining patterns of growth which are either normal or unusual. We use a functional version of principal components analysis to study the components of this variation, following closely the approach defined in Ramsay and Dalzell (1991). Our interest is primarily in the variation in the acceleration or the second derivative of height as a function of time. We are much less concerned about whether spline smoothing offers a better account of the actual data than kernel smoothing, certain parametric models, and other types of fit. Nevertheless, spline smoothing follows naturally from the way in which we formulate the problem of studying acceleration in height.

The data come from the Fels Growth Study, and an additional objective of the paper is to compare features observed in the Zurich data with those evident in these data.

2 The Data

The measurements of height used in this study span ages 2 to 22 years. The spacing between ages of observation vary from case to case, and only cases with around 40 observations and with intervals between measurements of 0.6 years or less were used. The sample sizes were 142 for boys and 116 for girls. For the convenience of the analyses to be described below, heights for the 41 half-yearly ages were estimated by linear interpolation and used in all subsequent calculations.

Figure 1 displays the averages and standard deviations for height for these ages and the two genders. The presence of the pubertal growth spurt (PS) is clearly evident in both curves, as is the difference in its timing for males and females. For the average boy PS begins at 9.5, reaches maximum velocity at 13.5, and is over by 19 years, while for girls the ages are 8.5, 11.5, and 17.

Figure 1 also displays the average first and second divided differences, $\Delta y_j = 2(y_j - y_{j-1})$ and $\Delta^2 y_j = 4(y_{j+1} - 2y_j + y_{j-1})$, as well as their standard deviations. In spite of the fact that the age of onset of PS is quite variable, it is clearly visible in these crude measures of velocity and acceleration. The same is not true, however, for the MS, which has been observed by Gasser, et al (1985) to be less variable in its location and therefore more likely to show up clearly in an average. If the MS were a feature present in some records and not others, the standard deviation should increase for these ages, but this fails to appear. There is some sign, however, of a brief deceleration preceding the onset of PS.

3 Splines and Functional Representation of the Data

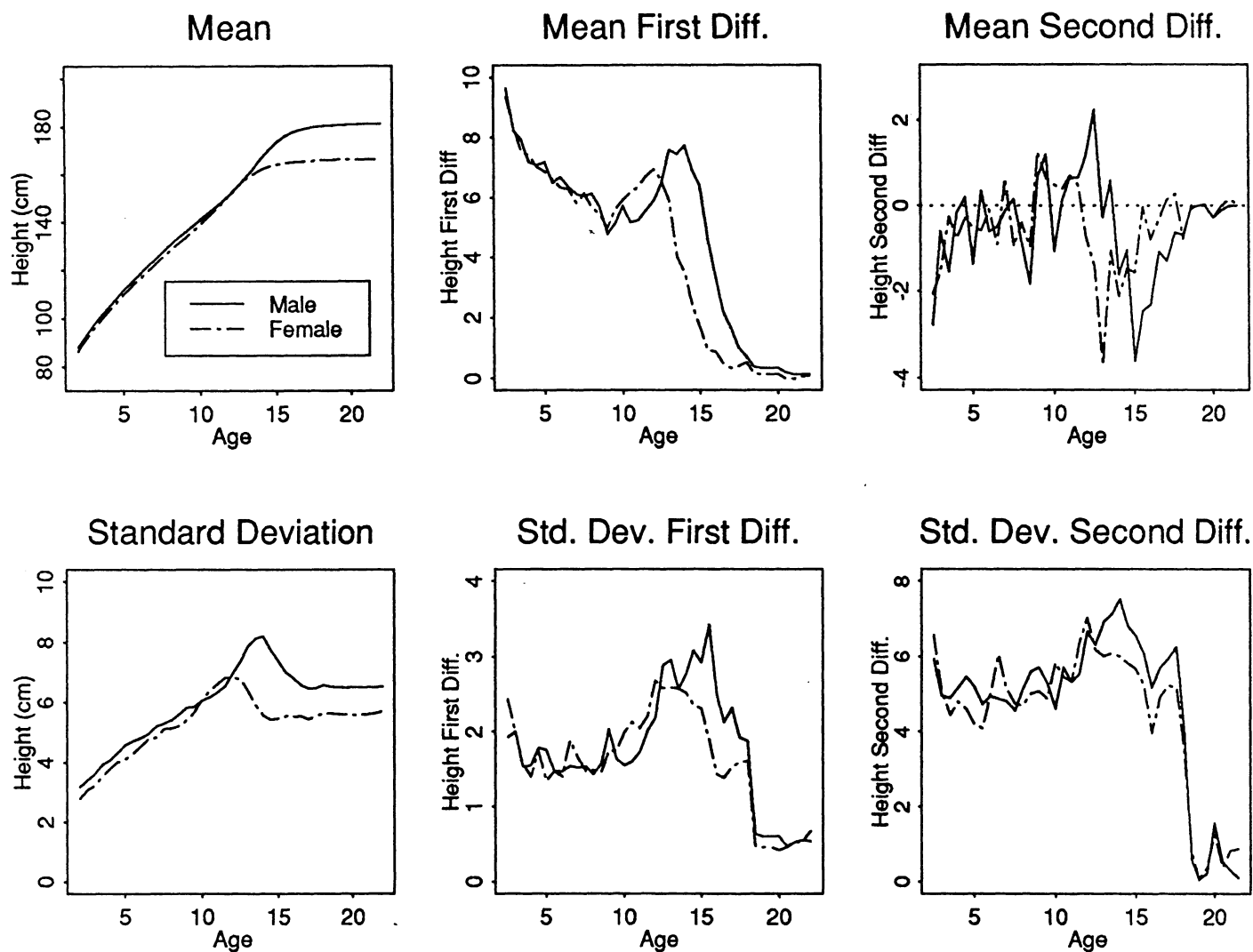
In order to be able to discuss functional concepts such as the behavior of the second derivative, we must first represent these data as a sample of smooth functions of time. The representation must take into account a small amount of error variation of the order of 2 mm. in any observation due a variety of factors such as the time of day at which the measurement was taken. Since we are specifically interested in the behavior of the second derivative, we shall want the functions representing the data to be conservative with respect to variation in the second derivative; that is, we don't want to see any more spurious variation in acceleration than we have to, and we hope to keep the signal-to-noise ratio high.

This leads naturally to the following smoothing criterion:

$$F(y|\lambda) = \min_h \{ (y - h)' \Sigma^{-1} (y - h) + \lambda \int (D^2 h)^2(t) dt \} \quad (2)$$

where h is the function which will represent the data, y is the vector of observed heights $(y_1, \dots, y_{41})'$, $h = (h(t_1), \dots, h(t_n))'$ is the vector of approximations to observed heights using the representing function h , residuals are assumed to have an order 41 covariance matrix Σ , $D^2 h$ is the the second derivative of h , and λ is a parameter which controls the smoothing process. Thus, the criterion to be minimized among candidate functions h consists of two terms, the first of which measures the lack of fit to the data given the

Figure 1: Means and standard deviations of height, and of its first and second divided differences.



residual covariance Σ , and the second of which is proportional to the size of the second derivative of h integrated over time.

The key to successful smoothing is to make λ just small enough so that the errors $y_j - h(t_j)$ begin to behave randomly, meaning typically normally with mean zero and covariance Σ . Various automatic techniques are available to determine an appropriate value of λ from the data themselves. A data-driven choice of λ can be made by minimizing the generalized cross-validation criterion GCV (Wahba, 1990) after estimation of the residual covariance Σ . This approach produced a value of λ of about 0.001, but after inspection of the fits and residuals, we decided to use $\lambda = 0.01$ for the results displayed in this paper. All analyses reported here were also repeated for a range of smoothing parameters.

The penalty term in (2) arises from the following inner product for functions:

$$\langle e, f \rangle = \int (D^2 e)(t)(D^2 f)(t) dt. \quad (3)$$

This inner product can also be used to define functional principal components analysis. In practical terms, we assess the agreement between two functions by integrating the product of their accelerations over time. One notes, however, that there is a potential problem with this measure of agreement, since adding a linear function $a + bt$ to either e or f will not change the inner product due to the fact that the second derivative of this component will vanish. Moreover, the inner product using (3) of any linear function and anything else will be automatically zero.

We circumvent this technical difficulty by confining our attention to functions that have no linear trend in a specified sense. We are free to choose this sense as we wish, and, for example, much of the theory of differential equations involves functions for which $h(0) = (Dh)(0) = 0$. We need a more natural definition of absence of linear trend, and will instead use the two constraints:

$$\int f(t) dt = 0 \quad \text{and} \quad \int t f(t) dt = 0 \quad (4)$$

Of course, such functions cannot account for the data themselves, and our actual smoothing function will have to contain a linear component, so that $h(t) = a + bt + f(t)$, but only $f(t)$ matters since we are working only with second derivatives. We can think of component $f(t)$ as the “de-tilted” or “de-trended” representation of the data.

In Figure 2 the de-trended component f , the first derivative of the latter, Df , and the second derivative $D^2h = D^2f$ are presented for the average boy and average girl. Comparing these curves with those in Figure 1, we see that f is the result of tilting and centering h so as to satisfy constraints (4). In the velocity curve Dh PS shows up as a local maximum, while for D^2 PS is indicated by a peak in acceleration at the onset, a crossing of the origin as linear growth is attained in the middle of PS, and finally a minimum as deceleration occurs at the end of PS and prior to a return to what is finally zero growth.

It can be shown (Ramsay and Dalzell, 1991) that the component f of a smoothing spline h satisfying constraints (4) has the form

$$f_i(t) = \sum_{j=1}^{41} c_{ij} k(t_j, t)$$

where the basis functions $k(t_j, t), j = 1, \dots, 41$, are

$$\begin{aligned} k_2(s, t) = [4 - 22(s + t) + 156st - 210st^2 + 70t^3 - 70(t^4 + s^4) \\ + 105st(s^3 + t^3) + 21(s^5 + t^5) - 42st(s^4 + t^4)]/420 \end{aligned} \quad (5)$$

$s \leq t$,

and $k_2(t, s) = k_2(s, t)$. This function is called the *reproducing kernel* for the Hilbert space of functions associated with the inner product (3).

Moreover, the coefficients c_{ij} are themselves linear combinations of the observed values y_j , so that we may express component $f(t)$ as

$$f(t) = \mathbf{k}_2(t)' \mathbf{A} \mathbf{y}$$

where \mathbf{k}_2 is the vector of functions $k_2(t_j, t), j = 1, \dots, 41$ and \mathbf{A} is a constant coefficient matrix. The sampling behavior of f , or any vector of its values, is therefore easy to work out in terms of the variation in \mathbf{y} , and in particular,

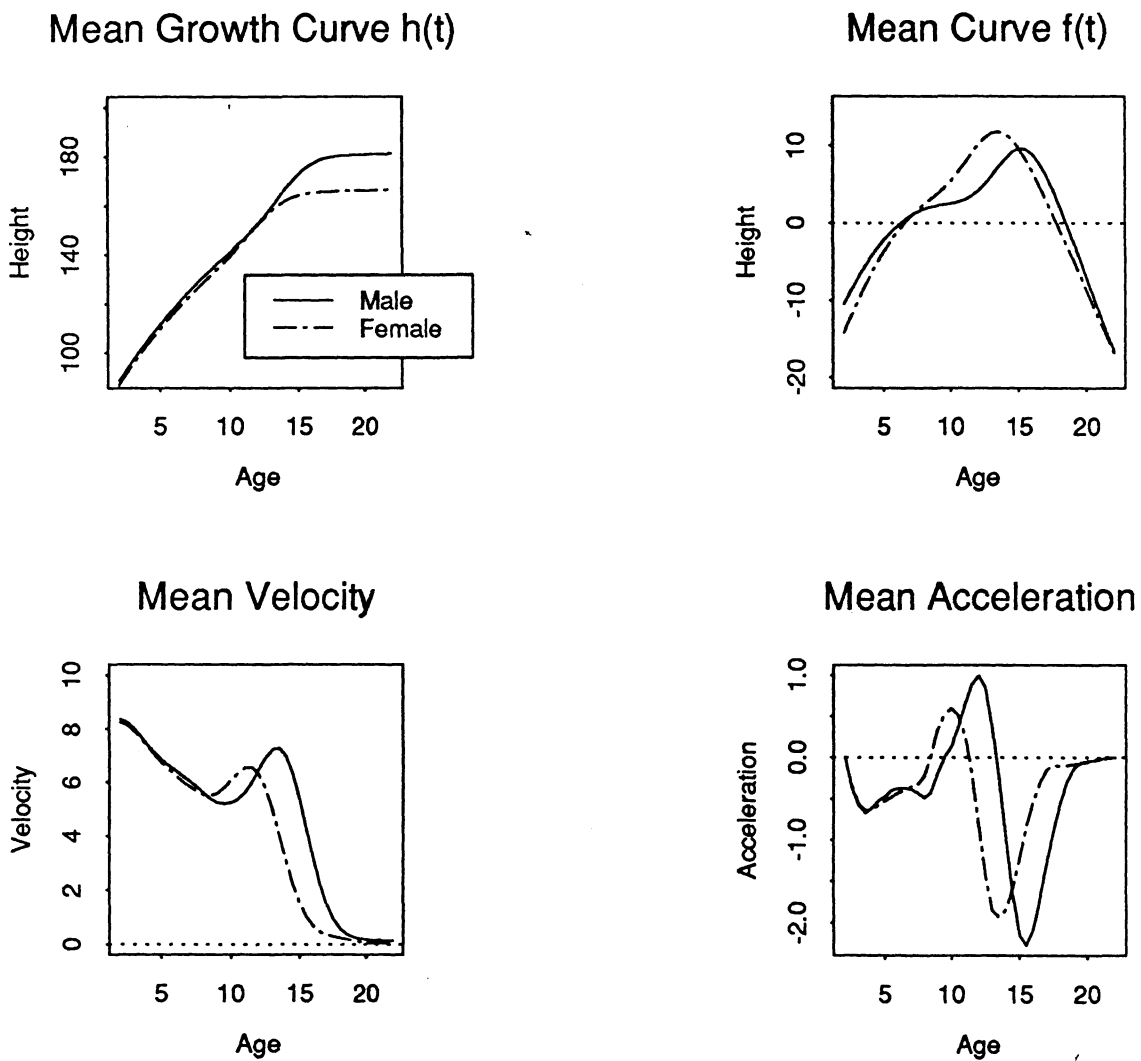
$$\text{Var}[f(t)] = \mathbf{k}_2(t)' \mathbf{A} \text{Var}(\mathbf{y}) \mathbf{A}' \mathbf{k}_2(t) \quad (6)$$

and

$$\text{Var}[D^2 f(t)] = D^2 \mathbf{k}_2(t)' \mathbf{A} \text{Var}(\mathbf{y}) \mathbf{A}' D^2 \mathbf{k}_2(t).$$

The availability of replications of the curves makes possible the estimation of $\text{Var}(\mathbf{y}) = \Sigma$ by the usual procedures, and one finds that the standard

Figure 2: Curves, detilted curves, first, and second derivatives for average males and females. The detilted curve $f(t)$ is the growth curve less the linear trend defined by constraints (3).



error of the estimated acceleration curve over pre-pubertal years but not less than five years is about $0.4cm^2$ per year. Consequently we can provide approximate point-wise confidence limits for curve locations. Figure 3 shows the estimated velocity and acceleration curves for the first male, who was fairly representative. Standard error estimates for kernel smooth estimates are rather similar.

4 Functional Principal Components

Instead of the original data, we are now in a position to work with the representing functions for males, $h_{iM}, i = 1, \dots, N_M$, and for female, $h_{iF}, i = 1, \dots, N_F$.

Defining PCA in functional terms is a matter of expressing the problem in inner-product or Hilbert space notation, which can serve equally well when the observations are either n p -vectors or n functions. In this system, PCA is the decomposition of

$$\max_{\langle \xi, \xi \rangle = 1} \{ \langle \xi, V\xi \rangle \} \quad (7)$$

where V is a self-adjoint linear operator which is, in the p -vector situation, the variance-covariance matrix, and in the functional data situation is the analogous bivariate variance-covariance kernel,

$$v(s, t) = n^{-1} \sum_{i=1}^n (h_i(s) - \bar{h}(s))(h_i(t) - \bar{h}(t)) .$$

The linear operator V is decomposed into principal components by finding the solutions of the eigenequation

$$\langle v(s, \cdot), \xi \rangle = \nu \xi(s) = \int (D^2 v(s, \cdot))(t) (D^2 \xi)(t) dt . \quad (8)$$

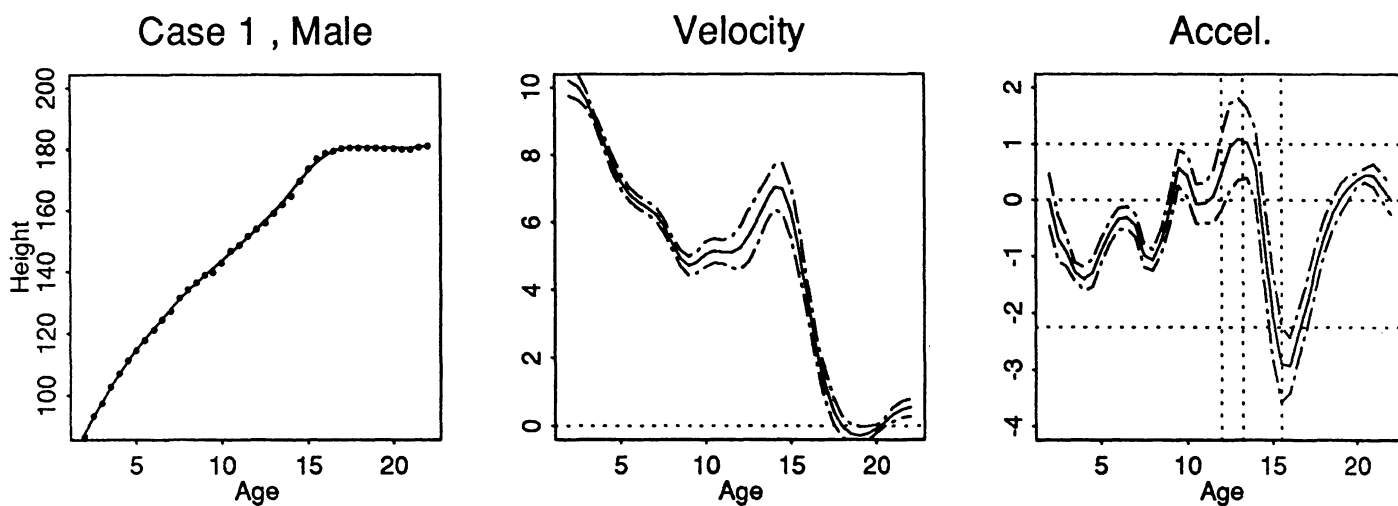
It follows (Besse and Ramsay, 1986) that the functional eigenequation (8) is equivalent to the matrix eigenequation

$$\mathbf{V}u = \nu u , \quad u' \mathbf{M} u = \mathbf{I} \quad (9)$$

where the *metric* \mathbf{M} is given by

$$\mathbf{M} = (\mathbf{K}_1 + \mathbf{K}_2)^{-1} \mathbf{K}_2 (\mathbf{K}_1 + \mathbf{K}_2)^{-1}$$

Figure 3: The velocity and acceleration functions for the first male. Dashed lines indicate point-wise 95% confidence limits for the function values. The vertical dotted lines for acceleration indicate the onset, peak velocity, and offset ages for the pubertal growth spurt for the average male.



where \mathbf{K}_2 is the order p matrix resulting from evaluating the reproducing kernel function k_2 at time points $t_j, j = 1, \dots, p$ and \mathbf{K}_1 is the corresponding matrix for the reproducing kernel $k_1(s, t)$ for linear functions of the form $u(t) = a + bt$.

Thus, we can ultimately return to multivariate analyses of the data, but with a change of metric which is of critical importance. Moreover, we can work backwards from this vector-valued PCA to obtain the corresponding results for the functional PCA, so that, for example, we can compute the eigenfunctions ξ orthonormal in the sense (3) once we have the \mathbf{M} -orthonormal eigenvectors u . Besse and Ramsay (1986) and Ramsay and Dalzell (1991) can be consulted for further technical details.

We chose to concentrate our attention on the first three eigenfunctions, which we like to refer to as *harmonics* as a tie-in with stochastic process theory. These accounted for 92.0% and 89.3% of the variance for males and females, respectively.

Figure 4 shows the three harmonics for both genders in the top two graphs. These do not appear to be orthogonal in the usual sense because their orthogonality is actually at the level of their second derivatives, and thus the bottom two displays graph the functions $D^2\xi$. When comparing the harmonics themselves, one must recall that variation in linear growth (overall height and overall rate) have already been eliminated by constraints (4).

Before attempting an interpretation of these harmonics, we should also have at our disposal the distribution of contributions of these harmonics to the actual growth curves. These we can calculate as the principal component scores $s_{im} = \langle h_i, \xi_m \rangle$, which likewise can be actually computed in terms of the eigenvectors u_m and the metric \mathbf{M} . These scores are in effect the Fourier coefficients for decomposing the observed curves h_i into orthonormal components ξ_m .

Figure 5 shows the variation in the first two scores. Also shown is a segment of an ellipse defined by the points (s_1, s_2) of the form

$$\begin{aligned} s_1(\theta) &= a_1 \cos(\theta) + c_1 \\ s_2(\theta) &= a_2 \sin(\theta) + c_2 \end{aligned} \tag{10}$$

where the constants a_1, a_2, c_1 , and c_2 have been estimated by least squares. The principal component scores show a strong tendency to fall along this locus.

Figure 4: The top two figures show the first three eigenfunctions ξ or harmonics for males and females, while the bottom two figures show their second derivatives $D^2\xi$.

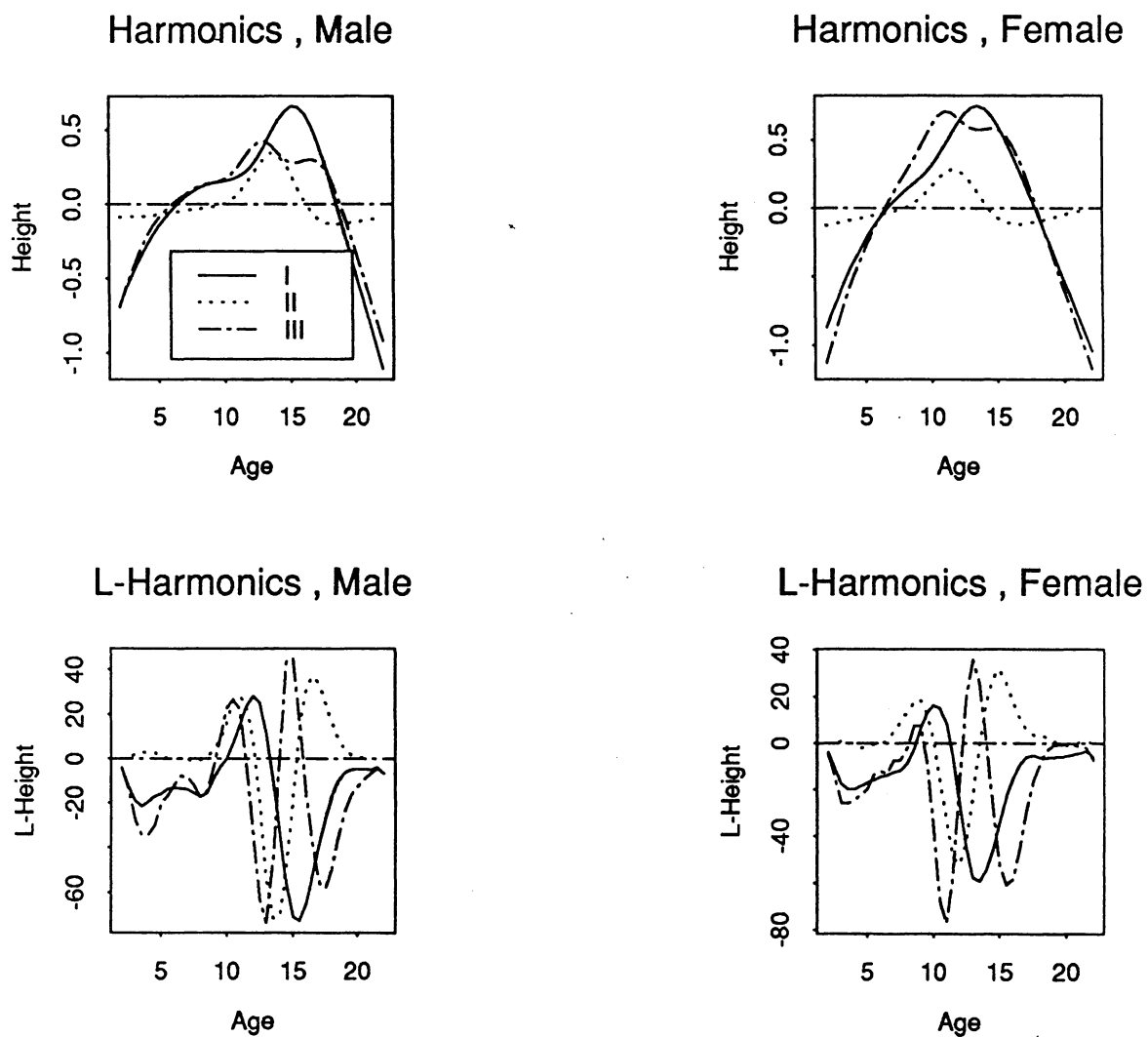
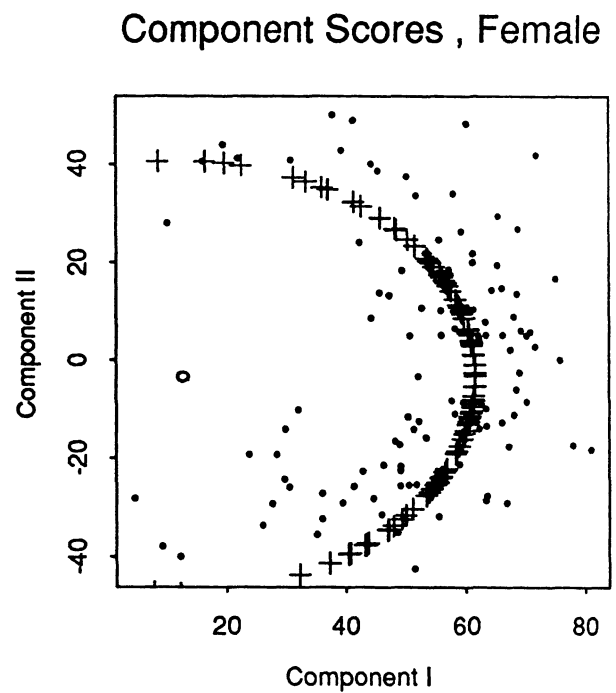
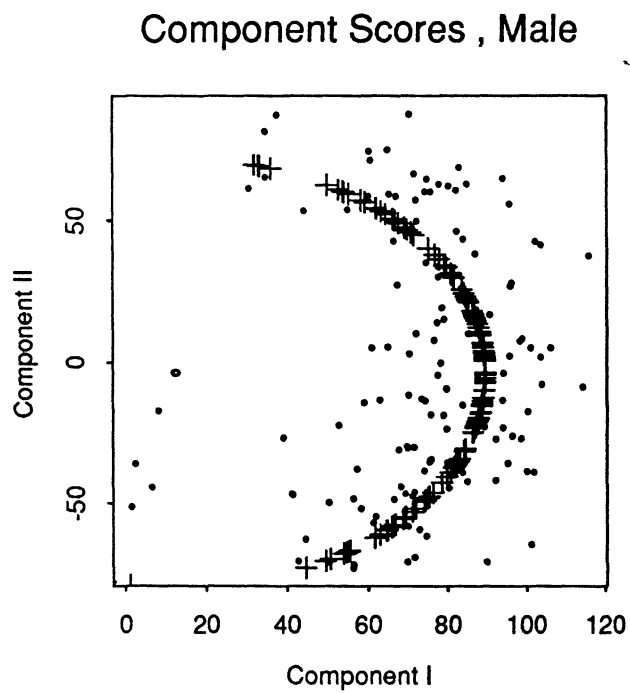


Figure 5: The solid dots indicate the principal components scores for the first two principal components of variation in acceleration. The crosses show the corresponding best fitting points on an ellipse, for which the center is given by the open circle.



The meaning of the angle $\theta_i = \tan^{-1}[(s_{i2} - c_2)/(s_{i1} - c_1)]$ is evident in Figure 6. The middle of the PS was estimated by estimating the age at which the acceleration curve changes sign as it goes from its positive peak to its negative trough. This is the age of maximum first derivative or velocity. We observe a very strong relationship between the angular measure of the first two principal component scores and this peak velocity age, suggesting that θ essentially measures the phase or the timing of the PS.

Similarly, Figure 7 indicates that the radial difference between the location of a point in Figure 5 and the fitted ellipse is determined primarily by the difference between the peak value of acceleration and the trough value, or the amplitude of the acceleration variation during PS. We might think of this as PS “force” or “energy”.

The only curves with strong contributions from the third harmonic are on the ends of the “horseshoe” pattern evident in these scores in Figure 5. The few cases on the anti-clockwise high-angle end of the ellipse are negative with respect to the third principal component $m = 3$, while cases on the clockwise low-angle end have positive scores on this component. Thus, we have a horseshoe pattern with one end bent down and the other bent up, but with most of the horseshoe relatively flat. It would appear that the third component essentially accommodates extreme phase values which the first two are unable to entirely model.

5 Is There a Mid-spurt in the Fels Data?

In most of the displayed examples in Gasser, et al (1984, 1985), the estimated acceleration curve exhibits a local maximum between six and eight years, but does not rise significantly above zero. Thus, the MS appears to differ qualitatively from the PS in being primarily a temporary cessation of linear growth rather than a period of accelerated growth. The data for the first male shown in Figure 3 show a MS fairly typical of those observed consistently in the Zurich data, although they also show a atypical secondary peak in acceleration at about nine years prior to the main PS peak.

The detection of a feature as small as this in the acceleration curves presents a number of problems. The temporal extent of the MS is well within the range of ages used to estimate the second derivative at any point. Both kernel smoothing and spline estimates of acceleration place substantial weight

Figure 6: The relation between the angle of a point in Figure 6 and the age at which the acceleration curve changes sign, or the midpoint of the PS.

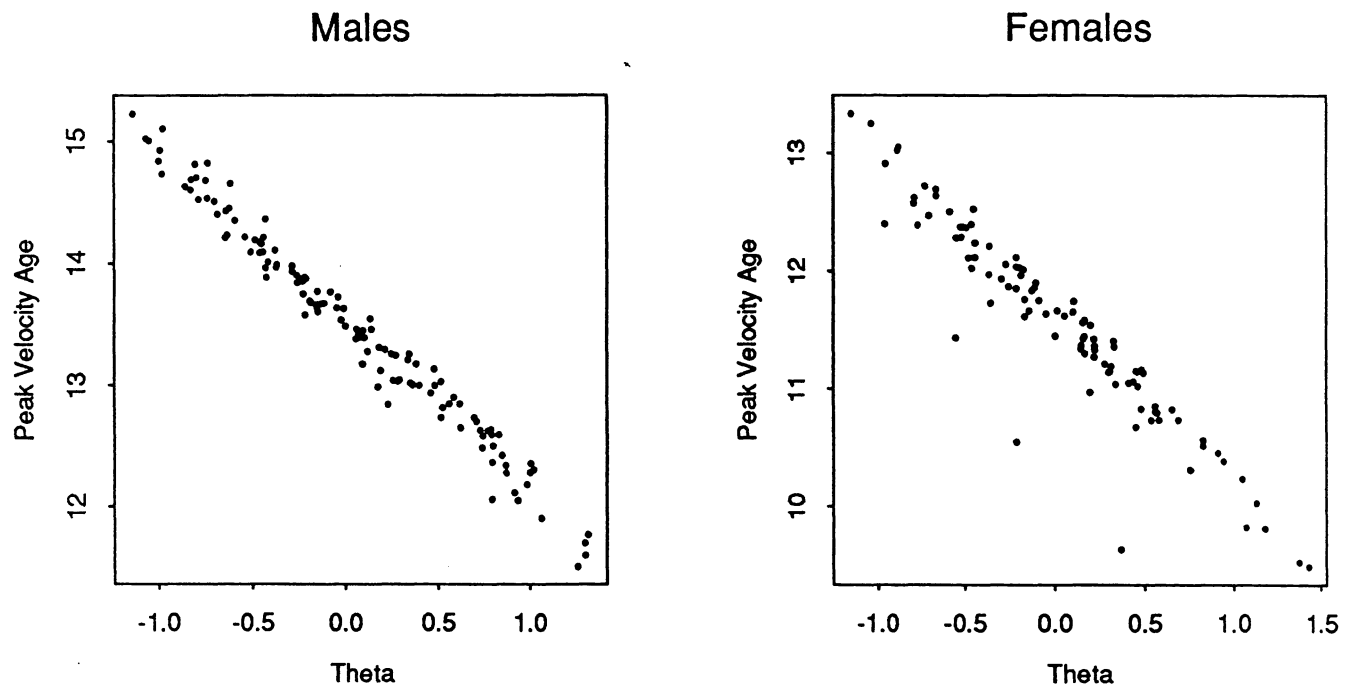
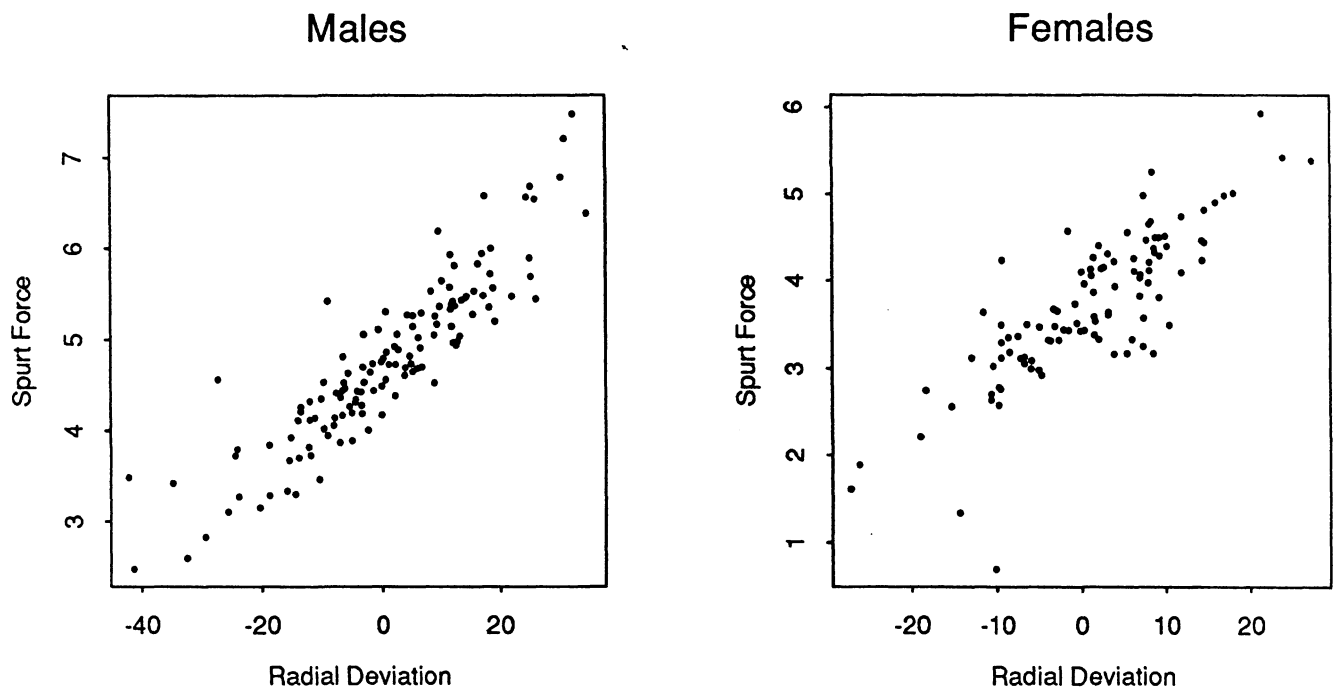


Figure 7: The radial discrepancy between a point in Figure 6 and the fitted ellipse is plotted against the change in acceleration from the maximum value prior to the mid-PS age and the following minimum.



on height observations three and even four years above and below the point of estimation. Figure 8 shows the weights used for the linear combination of height observations that estimates the acceleration at seven years.

The shape of the kernel implies a very strong autocorrelation between estimated accelerations with small lags, and Figure 8 shows that acceleration estimates separated by about two years can be expected to have a sampling correlation of around -0.4. This will tend to make estimated acceleration curves have “waves” of approximately the width of the MS, even when no actual spurt is present.

A third problem is presented by outlying height observations and long-tailed error distributions. A normal quantile plot of residuals for each point of observations revealed that, while many ages showed no real departure from Gaussian tail behavior, other ages indicated clear evidence of contamination of a basic Gaussian error process by one with a longer tail. Moreover, ages presenting such contamination showed no particular relationship to either gender or to proximity to features such as the PS and MS. This might arise, for example, if groups of children at a fixed age were measured in circumstances where less care was taken. Since both kernel and spline smooths are linear in nature, and since the weights vary greatly, the smooths will necessarily be fairly sensitive to large residuals.

Finally, even with Gaussian error, the pointwise confidence limits for estimated acceleration are fairly wide, as indicated in Figure 3. Simultaneous confidence limits will be even wider.

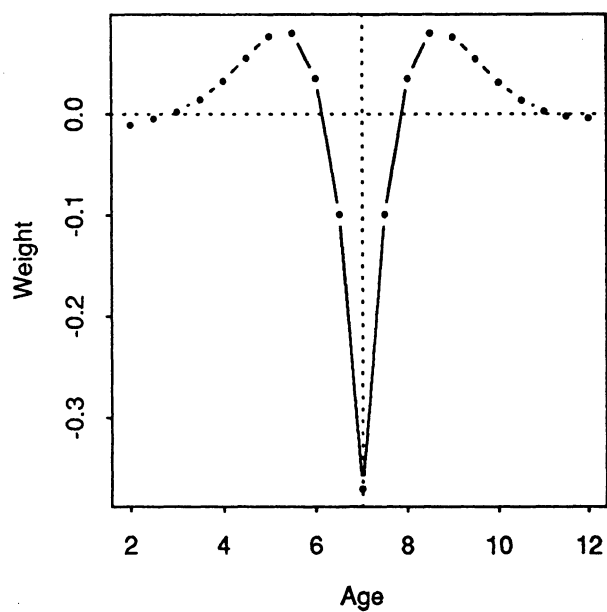
Our experiments with varying levels of smoothing, and with the use of various kernel smoothing estimates of acceleration, showed no consistent appearances of the midspurt for the Fels data. For illustration, Figure 9 shows ten randomly selected acceleration curves for males over the ages 5-9. None of these curves appears to clearly show a MS. Moreover, the t-ratios associated with the acceleration contrast $D^2h(8) + D^2h(6) - 2D^2h(7)$ were -0.02 for both genders.

The principal components harmonics displayed in Figure 4 also do not show any sign of MS. A principal components analysis of the accelerations over the age range 2 to 11 years also produced only one substantial eigenvalue, and the associated harmonic not surprisingly closely resembled the mean acceleration curve displayed in Figure 2.

Our tentative conclusion is that the term “mid-spurt” is somewhat misleading, in that the event is a pair of deceleration episodes, one at six years

Figure 8: The left display shows the weights used to estimate acceleration at seven years of age using a smoothing parameter value $\lambda = 0.01$. The autocorrelation in the estimated acceleration curve is shown in the right display.

Smoothing Opr. for Acceleration



Autocorrelation in Estimate

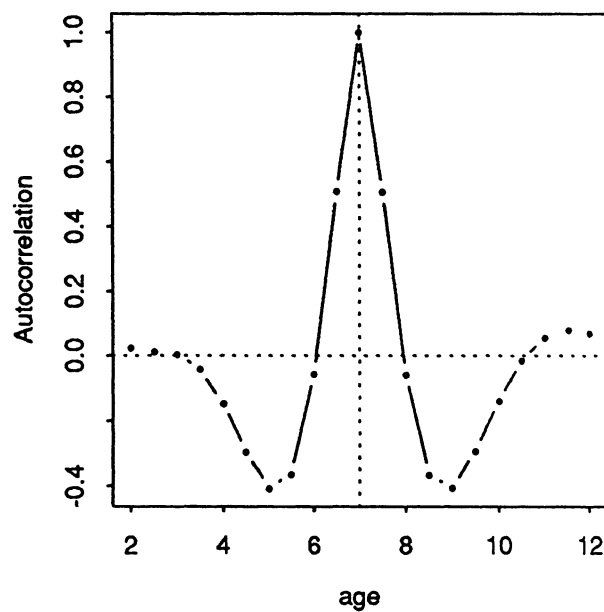
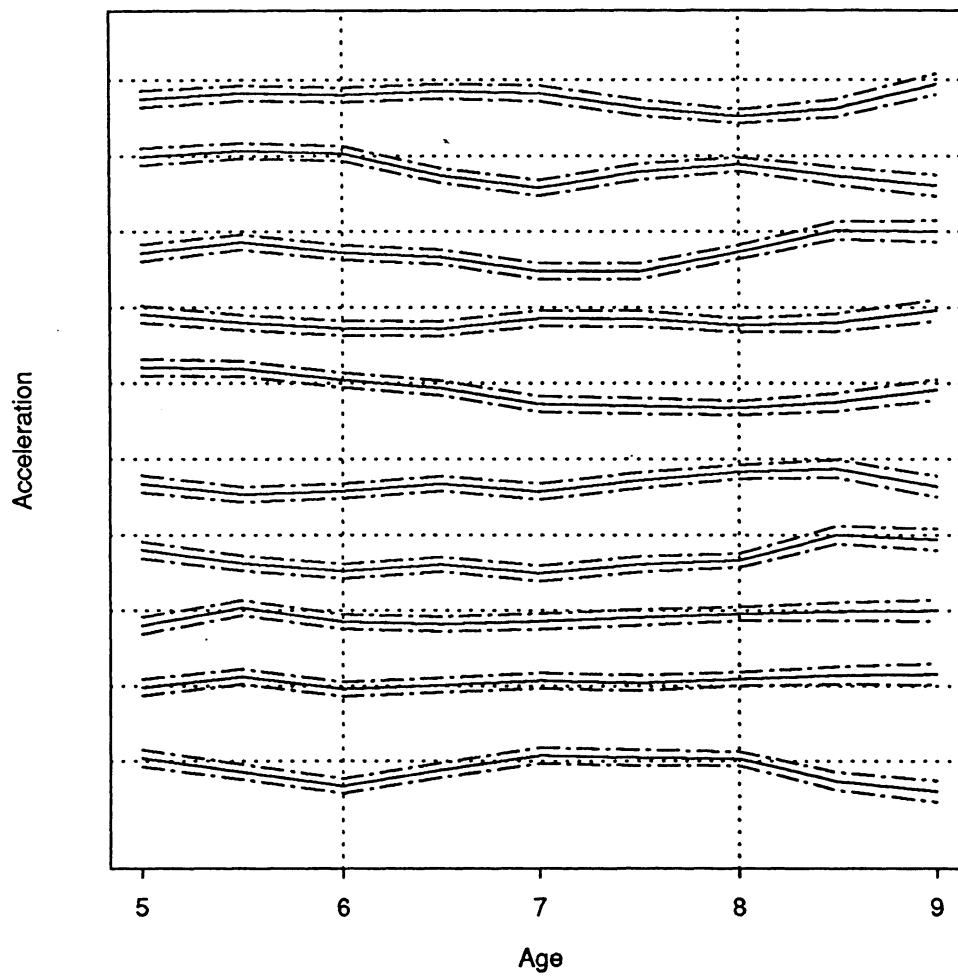


Figure 9: Ten randomly selected acceleration curves for the males (solid lines). Dashed lines surrounding each curve indicate pointwise 95% confidence limits, and the dotted line which is for the most part above each curve indicates zero acceleration.



and the other at about eight years, rather than a spurt. The evidence in the Zurich data for this event seems too strong to doubt, but we are at a loss to explain why the same phenomenon is not obvious in Fels data.

References

- Besse, P. and Ramsay, J. O. (1986) Principal components analysis of sampled functions. *Psychometrika*, 51, 285-311.
- Bock, R. D. and Thissen, D. (1980) Statistical problems of fitting individual growth curves. In F. E. Johnston, A. F. Roche and C. Susanne (Eds.) *Human Physical Growth and Maturation: Methodologies and Factors*. Plenum.
- Dauxois, J., Pousse, A., and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12, 136-154.
- Du Toit, S H. C. (1992) Multi-component models for non-consecutive data with ARMA(1,1) residuals. Unpublished manuscript, Department of Statistics, University of Pretoria.
- Eubank, R. L. (1988) *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Gasser, T., Köhler, W., Müller, H-G., Kneip, A., Largo, R., Molinari, L. and Prader, A. (1984) Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology*, 11, 397-411.
- Gasser, T., Müller, H-G., Köhler, W., Prader, A., Largo, R., and Molinari, L. (1985) An analysis of the mid-growth and adolescent spurts of height based on acceleration. *Annals of Human Biology*, 12, 129-148.
- Gasser, T., Kneip, A., Ziegler, P., Largo, R., and Prader, A. (1990) A method for determining the dynamics and intensity of average growth. *Annals of Human Biology*, 17, 459-474.
- Gasser, T., Kneip, A., Binding, A., Prader, A. and Molinari, L. (1991) The dynamics of linear growth in distance, velocity and acceleration. *Annals of Human Biology*, 18, 187-205.
- Gasser, T., Kneip, A., Ziegler, P., Largo, R., Molinari, L., and Prader, A. (1991) The dynamics of growth of weith in distance, velocity and acceleration. *Annals of Human Biology*, 18, 449-461.

- Jenss, R. M. and Bayley, N. (1937) A mathematical model for studying the growth of a child. *Human Biology*, 9, 556-563.
- Largo, R. H., Gasser, Th., Prader, A., Stützle, W., and Huber, P. J. (1978) Analysis of the adolescent growth spurt using smoothing spline functions. *Annals of Human Biology*, 5, 421-434.
- Preece, M. A. and Baines, M. J. (1978) A new family of mathematical models describing the human growth curve. *Annals of Human Biology*, 5, 1-24.
- Pütter, A. (1920) Studien über physiologische Ähnlichkeit. VI. Walkchstumsähnlichkeiten. *Pfluegers Archiv für die gesamte Physiologie*, 180, 298-340.
- Ramsay, J. O. and Dalzell, C. J. (1991) Some tools for functional data analysis (with discussion. *Journal of the Royal Statistical Society, Series B*, 53,, 539-572.
- Rao, C. R. (1987) Prediction in growth curve models (with discussion). *Statistical Science*, 2, 434-471.
- Rao, C. R. (1958) Some statistical methods for comparison of growth curves. *Biometrics*, 14, 1-17.
- Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.